On the Inconsistency of Instrumental Variables Estimators for the Coefficients of Certain Dummy Variables

David E. Giles

(Under revision, September, 2012)

web.uvic.ca/~dgiles/brown_bag/giles.pdf

THE PLAN

1. Some tricks with special dummy variables
2. Mention main results for OLS estimation
3. Extend results to I.V. estimation
4. Some simulation results
5. Conclusions

1. A very special dummy variable

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad ; \qquad \text{n observations}$$

$$\boldsymbol{d} = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{y} \\ y_{n+1} \end{pmatrix} = \begin{bmatrix} X & \boldsymbol{0} \\ \boldsymbol{x'}_{\boldsymbol{n+1}} & 1 \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \varepsilon_{n+1} \end{pmatrix} \qquad ; \quad \text{(n + 1) observations}$$

$$Q = \begin{bmatrix} X & \boldsymbol{0} \\ \boldsymbol{x'}_{\boldsymbol{n+1}} & 1 \end{bmatrix} \; ; \; Q' = \begin{bmatrix} X' & \boldsymbol{x}_{\boldsymbol{n+1}} \\ \boldsymbol{0}' & 1 \end{bmatrix}$$

$$Q'Q = \begin{bmatrix} X'X + \boldsymbol{x}_{\boldsymbol{n+1}}\boldsymbol{x'}_{\boldsymbol{n+1}} & \boldsymbol{x}_{\boldsymbol{n+1}} \\ \boldsymbol{x'}_{\boldsymbol{n+1}} & 1 \end{bmatrix}$$

Normal equations for OLS:

$$Q'Q\begin{pmatrix}\boldsymbol{\beta}\\ \gamma\end{pmatrix} = Q'\begin{pmatrix}\boldsymbol{y}\\ y_{n+1}\end{pmatrix}$$

or,

$$\begin{bmatrix}X'X + \boldsymbol{x}_{n+1}\boldsymbol{x'}_{n+1} & \boldsymbol{x}_{n+1}\\ \boldsymbol{x'}_{n+1} & 1\end{bmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \gamma\end{pmatrix} = \begin{pmatrix}X'\boldsymbol{y} + \boldsymbol{x'}_{n+1}y_{n+1}\\ y_{n+1}\end{pmatrix}$$

$$X'X\boldsymbol{\beta} + \boldsymbol{x}_{n+1}\boldsymbol{x'}_{n+1}\boldsymbol{\beta} + \boldsymbol{x}_{n+1}\gamma = X'\boldsymbol{y} + \boldsymbol{x'}_{n+1}y_{n+1} \qquad (1)$$

$$\boldsymbol{x'}_{n+1}\boldsymbol{\beta} + \gamma = y_{n+1} \qquad (2)$$

From (2): $\gamma = y_{n+1} - \boldsymbol{x'}_{n+1}\boldsymbol{\beta}$ (3)

Substitute in (1): $X'X\boldsymbol{\beta} = X'\boldsymbol{y}$

As usual, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$

Including “special” dummy variable is equivalent to dropping the observation.

From (3): $\hat{\gamma} = y_{n+1} - \boldsymbol{x'}_{\boldsymbol{n+1}}\widehat{\boldsymbol{\beta}}$

$$e_{n+1} = \left(y_{n+1} - \boldsymbol{x'}_{\boldsymbol{n+1}}\widehat{\boldsymbol{\beta}} - 1\,\hat{\gamma}\right) = 0\ .$$

Questions:

- Does this happen with other estimators?
- What if there are just 2 observations for which the dummy variable = 1?
- Why would we want to include such a dummy variable?
- What can we say about the standard error & t-statistic associated with $\hat{\gamma}$?

2. Other OLS results – Hendry & Santos (2005)

- $\hat{\beta}$ and $\hat{\gamma}$ are BLUE.
- $\hat{\beta}$ is weakly consistent for β.
- The t-statistic for testing $H_0 : \beta_j = \beta_j^*$, for the j^{th} element of β, is still Student-t distributed under the null.
- $\hat{\gamma}$ is an **inconsistent** estimator of γ.
- The usual estimator of the variance of $\hat{\gamma}$ is still unbiased and consistent.
- The t-test statistic for testing $H_0 : \gamma = \gamma *$ is still Student-t distributed under the null.
- The latter t-test is **inconsistent**.

3. Extension to I.V. (*this paper*)

$$y = X\beta + D\gamma + v \qquad ; \qquad v \sim [0, \sigma^2 I]$$

- At least some of the columns of *X* $(n \times k_1)$ are random and correlated even asymptotically with the error term, *v*. That is $plim(n^{-1}X'v) \neq 0$.

- The columns of *D* $(n \times k_2)$ are zero-one indicator variables, each taking the value unity only for one (different) observation.

- Without loss of generality, include the intercept and all of the columns of *D* in the set of instruments.

- Let the columns of *Z* $(n \times k_1)$ be the remaining k_1 instruments, satisfying $plim(n^{-1}Z'v) = 0$ and $plim(n^{-1}Z'X) = Q_{ZX}$, where Q_{ZX} is finite and non-singular.

For expository purposes, let $k_2 = 1$, so that *D* is a single column vector, *d*, with one non-zero element at observation, i_b, say.

- Note that $d'd = 1$, $X'd = x_{i_b}$ and $Z'd = z_{i_b}$ where x_{i_b} and z_{i_b} are $(k_1 \times 1)$ vectors with elements comprising the values of the regressors in *X* and the instruments in *Z*, respectively, at observation i_b.

- So, $plim(n^{-1}d'X) = 0'$.

- Also, define $M_1 = I - X(Z'X)^{-1}Z'$ and $M_2 = I - d(d'd)^{-1}d' = I - dd'$, so that $M_1X = 0$ and $M_2 d = 0$.

- Note that $plim(d'M_1 d) = 1 - plim[(n^{-1}d'X)(n^{-1}Z'X)^{-1}Z'd] = [1 - 0'Q_{ZX}^{-1} z_{i_b}] = 1.$

Applying an I.V. version of the Frisch-Waugh-Lovell Theorem (Giles, 1984) –

$$\tilde{\beta} = (Z'M_2X)^{-1}Z'M_2y \quad \text{and} \quad \tilde{\gamma} = (d'M_1d)^{-1}d'M_1y$$

Theorem 1

$\tilde{\gamma}$ is an inconsistent estimator of γ.

Proof

$$\begin{aligned}\tilde{\gamma} &= (d'M_1d)^{-1}d'M_1y \\ &= (d'M_1d)^{-1}d'M_1(X\beta + d\gamma + v) \quad .\end{aligned}$$

Using the result, $M_1X = 0$, we have:

$$\begin{aligned}\tilde{\gamma} - \gamma &= (d'M_1d)^{-1}d'M_1v \\ &= (d'M_1d)^{-1}d'v - (d'M_1d)^{-1}d'X(Z'X)^{-1}Z'v \quad .\end{aligned}$$

$$\tilde{\gamma} - \gamma = (d'M_1d)^{-1}d'v - (d'M_1d)^{-1}d'X(Z'X)^{-1}Z'v \quad .$$

By Slutsky's Theorem,

$$plim[(d'M_1d)^{-1}(d'X)(n^{-1}Z'X)^{-1}(n^{-1}Z'v)] = (x'_{i_b}\, Q_{ZX}^{-1}\, 0) = 0 \quad .$$

Also, recalling that $plim(d'M_1d) = 1$,

$$plim[(d'M_1d)^{-1}(d'v)] = plim(v_{i_b}) = v_{i_b} \quad ,$$

where v_{i_b} is a single element of *v* .

So, using Slutsky's Theorem again,

$$plim(\tilde{\gamma} - \gamma) = v_{i_b} \neq 0 \quad . \qquad \#$$

[If the dummy variable takes the value unity for a *fixed* number of observations (the first *m*, say), and this number does not increase with *n*, then $\tilde{\gamma}$ is still inconsistent.]

Theorem 2

$\tilde{\beta}$ is a consistent estimator of β.

Theorem 3

If $plim(n^{-1}X'\nu)$ is a finite vector; and $plim(n^{-1}X'X) = Q_{XX}$ and $plim(n^{-1}Z'Z) = Q_{ZZ}$ are finite matrices. Then the asymptotic variance of $\sqrt{n}(\tilde{\gamma} - \gamma - \nu_{i_b})$ is $(\sigma^2 x_{i_b}' Q_{ZX}^{-1} Q_{ZZ} Q'^{-1}_{ZX} x_{i_b})$; and this asymptotic variance can be estimated consistently by $n\tilde{\sigma}^2 x_{i_b}'(Z'X)^{-1}Z'Z(X'Z)^{-1}x_{i_b}$, where

$$\tilde{\sigma}^2 = (\tilde{\nu}'\tilde{\nu}/n) = (y - X\tilde{\beta} - d\tilde{\gamma})'(y - X\tilde{\beta} - d\tilde{\gamma})/n$$

is the usual consistent estimator of σ^2.

[This estimator of the covariance matrix is the usual one that we would construct.]

Theorem 4

Let $Z_{(i_b)}$ represent the Z matrix with the i_b th row deleted, and assume that $plim(n^{-1}Z_{(i_b)}'Z_{(i_b)}) = Q_{Z*Z*}$ is a finite matrix. Then the a.c.m. of $\sqrt{n}(\tilde{\beta} - \beta)$ is $(\sigma^2 Q_{ZX}^{-1} Q_{Z*Z*} Q'^{-1}_{ZX})$; and this asymptotic variance can be estimated consistently by $n\tilde{\sigma}^2 x_{i_b}'(Z'X)^{-1} Z_{(i_b)}'Z_{(i_b)}(X'Z)^{-1}$, where

$$\tilde{\sigma}^2 = (\tilde{v}'\tilde{v}/n) = (y - X\tilde{\beta} - d\tilde{\gamma})'(y - X\tilde{\beta} - d\tilde{\gamma})/n$$

is the usual consistent estimator of σ^2.

[This estimator of the covariance matrix is the usual one that we would construct.]

Theorem 5

The usual t-test statistic for testing $H_0 : \beta_j = \beta_j^*$, for the j^{th} element of β, is asymptotically standard normally distributed if the null hypothesis is true.

Theorem 6

The usual t-test statistic for testing $H_0 : \gamma = \gamma^*$ is asymptotically standard normally distributed if the null hypothesis is true.

[This follows from the asymptotic normality of $\tilde{\gamma}$ and the asymptotic independence of $(\tilde{\gamma} - \gamma^*)$ and its estimated asymptotic variance, $n\tilde{\sigma}^2 x_{i_b}'(Z'X)^{-1}Z'Z(X'Z)^{-1}x_{i_b}$. So, asymptotically valid inferences may still be drawn about the coefficients of dummy variables that take only a **fixed** number of non-zero values. Same for Wald test.]

Theorem 7

The usual t-test for testing $H_0 : \gamma = \gamma^*$ is inconsistent.

4. Some simulation results

$$x_i = a_0 + a_1 d_i + a_2 z_i + \varepsilon_i \qquad (4)$$

$$y_i = \alpha + \beta x_i + \gamma d_i + \nu_i \qquad (5)$$

$$\begin{pmatrix} \varepsilon_i \\ \nu_i \end{pmatrix} \sim N\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}; \qquad i = 1, 2, ..., n.$$

- Sample sizes up to n = 200,000 were considered.
- Exogenous variable, z, generated as standard normal for the largest sample size, and held fixed for the 1,000 Monte Carlo repetitions.
- Dummy variable, d: $d_i = 1$, i = 1, 2, ... , 15; $d_i = 0$, i = 16, 17, ... , n.

  $a_0 = a_1 = a_2 = \alpha = \beta = \gamma = \sigma_1^2 = \sigma_2^2 = 1$; $\sigma_{12} = 0.95$.
- Equation (5) was estimated by I.V. with d, z and the intercept as instruments.

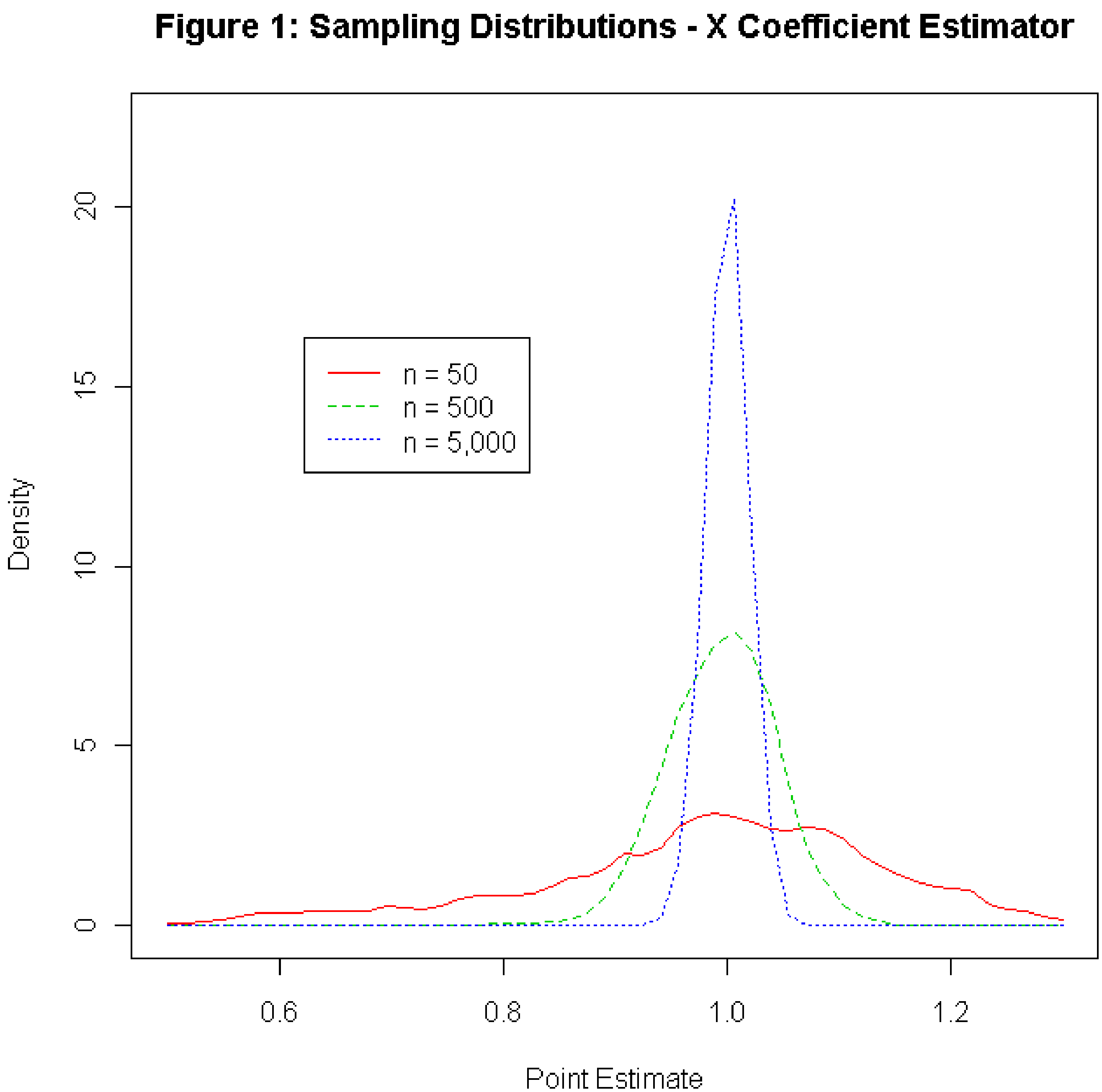


Figure 1: Sampling Distributions - X Coefficient Estimator
n = 50
n = 500
n = 5,000
Density
0
5
10
15
20
0.6
0.8
1.0
1.2
Point Estimate

Figure 2: Sampling Distributions - Dummy Coefficient Estimator

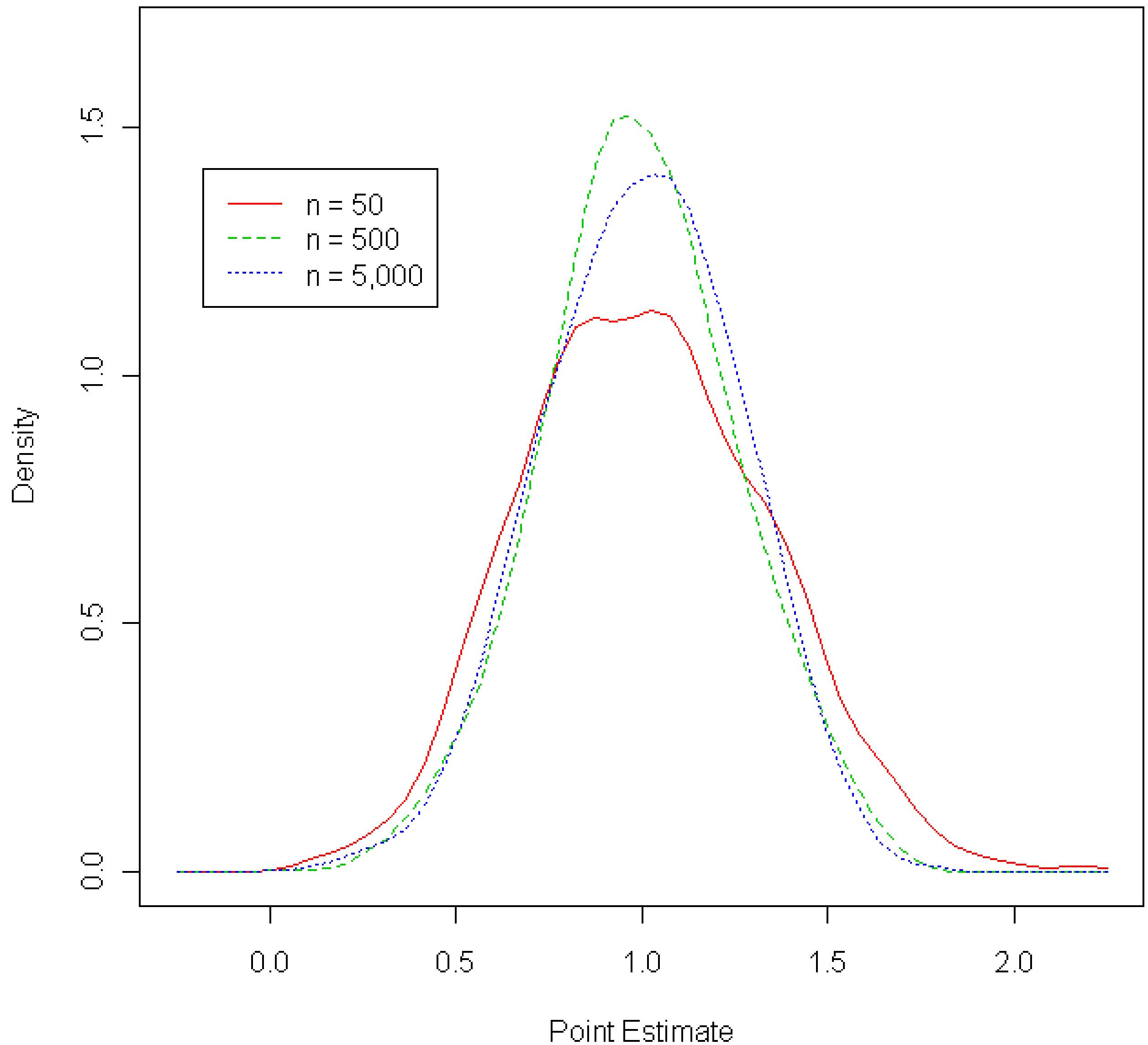
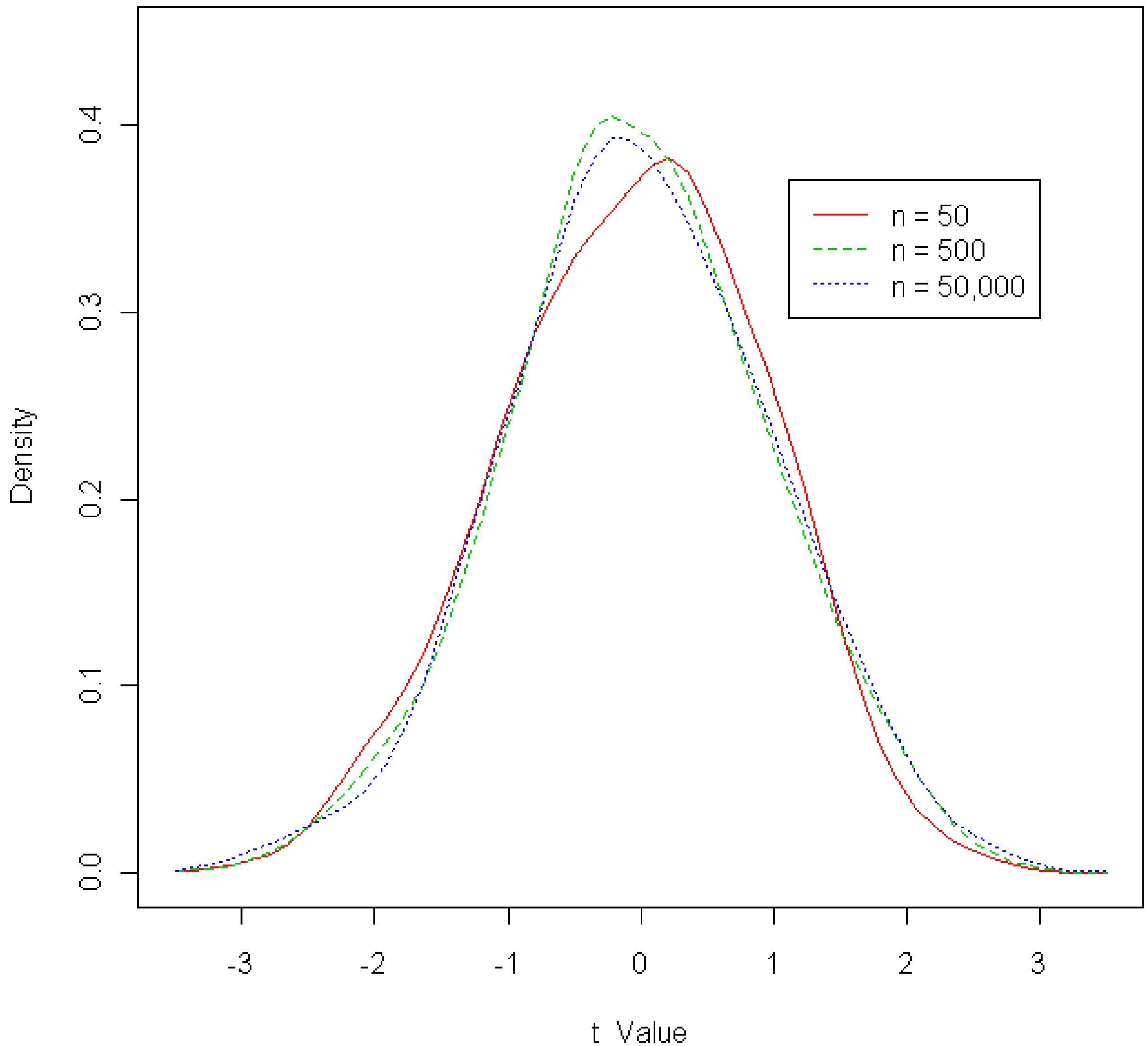


Figure 3: Sampling Distributions - Dummy Coefficient t-Statistic



	Mean	S.D.	Jarque-Bera p-value
n			
50	-0.053	0.979	0.020
500	-0.027	0.980	0.821
50,000	-0.008	1.007	0.806

Table 1: Powers of t-test of $H_0 : \gamma = 0$ *vs.* $H_A : \gamma \neq 0$

(Size = 5%)

	γ					
	-0.5	-0.3	-0.1	0.1	0.3	0.5
n						
50	0.37	0.19	0.06	0.05	0.12	0.31
500	0.49	0.21	0.07	0.06	0.20	0.46
5,000	0.49	0.21	0.07	0.07	0.21	0.48
100,000	0.49	0.22	0.07	0.07	0.22	0.49
200,000	0.49	0.22	0.07	0.07	0.22	0.49

6. Conclusions

- Simple “impulse” dummy variables can give rise to unusual asymptotics.
- The usual estimators of the coefficients of impulse dummies are inconsistent.
- However, the usual confidence intervals and tests of restrictions will still be valid, asymptotically.
- So, we can still test if outliers are “significant”.
- But, these tests are “inconsistent” – so their power is limited even for large “n”.
- Things get more complicated if we have a dynamic time-series model
- Things also get more complicated if the X data are non-stationary and/or cointegrated.